A note on stochastic volatility model estimation
(Nota sobre estimação de modelos de volatilidade estocástica)

Omar Abbara[†]
Mauricio Zevallos[‡]

Abstract

The paper assesses the method proposed by Shumway and Stoffer (2006, Chapter 6, Section 10) to estimate the parameters and volatility of stochastic volatility models. First, the paper presents a Monte Carlo evaluation of the parameter estimates considering several distributions for the perturbations in the observation equation. Second, the method is assessed empirically, through backtesting evaluation of VaR forecasts of the S&P 500 time series returns. In both analyses, the paper also evaluates the convenience of using the Fuller transformation.

***Keywords*:** Backtesting; Mixtures; Non-Gaussian errors; Value-at-Risk
JEL Code: C22, C58, G15.

1. Introduction

Stochastic volatility (SV) models, initially proposed by Taylor (1982, 1986), compose a well-known class of models to estimate volatility. These models have gained attention in the financial econometrics literature because of their flexibility to capture the nonlinear behavior observed in financial time series returns. In this sense, they are an attractive alternative to many ARCH-type models which are based on the seminal work of Engle (1982).

Let $r_1,\ldots,r_n$ be a time series of financial returns. In the SV model proposed by Taylor (1982, 1986),

$$r_t = \beta e^{h_t/2}\varepsilon_t, \tag{1}$$
$$h_t = \phi h_{t-1} + \omega_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent identically distributed (IID) random variables with zero mean and unit variance, $\{\omega_t\}$ is an IID sequence with $\omega_t \sim N(0,\sigma_\omega^2)$, and $\{\varepsilon_t\}$ and $\{\omega_t\}$ are assumed to be independent for all t. In this model the volatility at time t, given by $\sigma_t = \beta\exp(h_t/2)$, is unobserved because h_t is a latent variable.

To estimate model (1)-(2), several methods have been proposed in the literature, many of them based on the strategy of taking the log of the squared returns in (1) to obtain

$$y_t = \alpha + h_t + \eta_t, \tag{3}$$

where $y_t = \ln(r_t^2)$, $\alpha = \ln(\beta^2)$ and $\eta_t = \ln(\varepsilon_t^2)$. Then (3)-(2) is a state space model with non-Gaussian perturbations η_t. Therefore the challenge is how to perform inference in a non-Gaussian framework. Harvey, Ruiz & Shephard (1994) considered a quasi-maximum likelihood approach where perturbations η_t are assumed Gaussian and estimates are obtained by maximizing the likelihood constructed via the Kalman filter. However, this method can produce estimates with high bias and variance when the variability of the latent process is small relative to the variability of the log-squared returns. To improve the efficiency of the latter proposal, Kim, Shephard & Chib (1998) proposed a Markov Chain Monte Carlo (MCMC) method where the distribution of η_t is approximated by a mixture of seven normals with probabilities, means and variances previously specified. For an account of other estimation methods including MCMC alternatives, see the reviews of Broto & Ruiz (2004) and Shephard & Andersen (2009).

Submitted on July 29, 2019. Revised on December 10, 2019. Accepted on December 10, 2019. Published online in December 2019. Editor in charge: Marcelo Fernandes.
[†]Canvas Capital S.A., Brazil. E-mail: muhieddine@gmail.com
[‡]Department of Statistics, University of Campinas, Brazil. E-mail: amadeus@unicamp.br

©2019 Sociedade Brasileira de Finanças, under a Creative Commons Attribution 3.0 license (http://creativecommons.org/licenses/by/3.0)

An interesting method for maximum likelihood estimation was proposed by Shumway & Stoffer (2006, Chapter 6, Section 10). They also approximated η_t as a mixture of normals:

$$\eta_t = \sum_{i=1}^{m} I_{t,i} V_{t,i}, \tag{4}$$

where $\{V_{t,i}\}$ is an IID sequence with $V_{t,i} \sim N(\mu_i, s_i^2)$, $\mu_1 = 0$, and $\{I_{t,i}\}$ is an IID Bernoulli sequence with $P[I_{t,i} = 1] = \pi_i$ and $\pi_i = 1/m$. Then, adopting the approximation given by Equation (4), the state space model (3)-(2) is conditionally Gaussian. Therefore it is possible to obtain the innovations via a Kalman filter algorithm and write the likelihood function based on these innovations, as follows. Let $h_{t|t-1} = E(h_t|y_1,...,y_{t-1})$ be the predicted values of h_t given the past information; the Kalman recursions are given by:

$$h_{t+1|t} = \phi h_{t|t-1} + \sum_{j=1}^{m} \pi_{tj} k_{tj} \varepsilon_{tj}, \tag{5}$$

$$P_{t+1|t} = \phi^2 P_{t|t-1} + \sigma_\omega^2 - \sum_{j=1}^{m} \pi_{tj} k_{tj}^2 \varepsilon_{tj}, \tag{6}$$

$$\varepsilon_{tj} = y_t - \alpha - h_{t|t-1} - \mu_j, \tag{7}$$

$$\Sigma_{tj} = P_{t|t-1} + \sigma_j^2, \tag{8}$$

$$k_{tj} = \phi^2 P_{t|t-1} / \Sigma_{tj}, \tag{9}$$

where the probabilities $\pi_{jt} = P(I_{jt} = 1|y_1,...,y_t)$ are calculated by:

$$\pi_{jt} = \frac{\pi_j p_j(t|t-1)}{\sum_{j=1}^{m} \pi_j p_j(t|t-1)}, \tag{10}$$

with the density $p_j(t|t-1)$ being approximated by a $N(h_{t|t-1} + \mu_j, \Sigma_{jt})$. Then, the log-likelihood is evaluated by:

$$\ln L(\Theta) = \sum_{t=1}^{n} \ln \left(\sum_{j=1}^{m} \pi_j p_j(t|t-1) \right). \tag{11}$$

In order to compute expressions in (5)-(11), one must first specify the number of terms in the mixture, m, as well as some initial values. Shumway and Stoffer considered $m = 2$, $h_{1|0} = 0$ and $P_{1|0} = \phi^2 + \sigma_\omega^2$.

The estimation of parameters $\Theta = (\alpha, \phi, \sigma_\omega, \mu_1, \ldots \mu_m, \sigma_1, \ldots, \sigma_m)$ is obtained by maximizing the innovation-based likelihood function given by Equation (11). Note that unlike Kim at al. (1998), the parameters of the mixture given by μ_i and σ_i, for $i \in \{1,...,m\}$, have to be estimated.

As far as we know, the method proposed by Shumway and Stoffer has not been evaluated in terms of the quality of the point estimates. Thus, here we intend to fill this gap in the literature by assessing the performance of the estimation method through Monte Carlo simulations. Specifically, we evaluate the performance when ε_t in (1) follows different distribution functions, ranging from symmetric to asymmetric ones. There is empirical evidence of the convenience of using non-Gaussian distributions for ε_t, as discussed by Bai, Russell & Tiao (2003) and Liesenfeld & Jung (2000). We also study the behavior of the estimates when the mixture in (4) is defined with $m = 2$ and $m = 3$ terms. As mentioned above, Shumway and Stoffer considered $m = 2$ and here our objective is to assess if the inclusion of a third term in the mixture improves the estimation even at the cost of loss of precision (because the number of parameters increases).[1] In addition, we evaluate the convenience

[1] A referee expressed concerns about the use of only $m = 3$ in the mixture approximation given the results of Kim et al. (1998) who used $m = 7$ to approximate the distribution of the logarithm of a log chi-square distribution with one degree of freedom. However, the quality of the estimates of the parameters (as well as the volatility estimates) also and crucially depend on factors such as the filtering algorithm, the convexity of the log-likelihood and the optimization routine. In the proposal of Shumway and Stoffer, one has to estimate two parameters for each extra term in the mixture, implying less precision and possibly non-convergence of parameters (especially in small samples like that used in the Monte Carlo experiments of the paper). In addition, the approximation proposed by Kim et al. (1998) only holds for Gaussian perturbations and, as far as we know, the case of mixture approximations for non-Gaussian perturbations (very important in applications) has not been studied.

of using the Fuller transformation of the data. This transformation, suggested by Fuller (1996, pp.494–497), is denoted by:

$$y_t = \ln(r_t^2 + cs^2) - \frac{cs^2}{r_t^2 + cs^2}, \tag{12}$$

where s^2 is the sample variance of r_t and c is a small constant, used to avoid zero returns (which leads to non-finite values of log squared returns) or mitigate the effects of inliers. In short, our main goal of the paper is to provide practitioners with valuable clues to apply the method of Shumway and Stoffer. On the other hand, the quality of volatility estimates is assessed empirically through a backtesting exercise of one-step-ahead Value-at-Risk (VaR) forecasts of the S&P 500 time series returns. The VaR calculations were performed by a novel procedure and this topic was not addressed in the proposal of Shumway and Stoffer.

The remainder of the paper is organized as follows. In Section 2 we present the results of the Monte Carlo experiments; Section 3 presents the results of the backtesting exercise of VaR forecasts of S&P 500 returns; and Section 4 summarizes the main conclusions.

2. Monte Carlo experiments

2.1 Setup

For each combination of parameters, we simulated 2,000 replications of time series of size $n = 1{,}250$ from model (1)-(2), assuming that ε_t disturbances follow the standard Gaussian distribution, two heavy-tailed distributions and three skew distributions. The heavy-tailed distributions are the t-Student distribution and the generalized error distribution (GED), with probability density functions (pdf) defined as:

$$t(\varepsilon_t|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\varepsilon_t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{13}$$

$$g(\varepsilon_t|\nu) = \frac{\nu}{\lambda 2^{1+1/\nu}\Gamma(\frac{1}{\nu})}\exp\left(-0.5\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu}\right), \tag{14}$$

respectively, where $\lambda = \sqrt{2^{-2/\nu}\Gamma\left(\frac{1}{\nu}\right)\Gamma\left(\frac{3}{\nu}\right)}$. The GED distribution whose pdf is given by Equation (14) has zero mean and unit variance and the t-Student distribution in (13) has to be standardized in order to have variance equal to 1. For those distributions, we considered parameters equal to $\nu = 5$ for the t-Student and $\nu = 1.5$ for the GED to reproduce heavy-tailed distributions. These values of ν imply kurtosis values around 4 and 9 for GED and t-Student distributions, respectively. Similar values of ν have been obtained in empirical applications; see Zevallos, Gasco & Ehlers (2017) when using SV models, and Nelson (1991) when using EGARCH models.

In addition, to generate ε_t with asymmetric distribution we used the method proposed by Fernandez & Steel (1998). Here, let $f(.)$ be a univariate unimodal probability density function symmetric at zero. Then, a skewed pdf, say $p(.)$, can defined through

$$p(\varepsilon_t|\gamma) = \frac{2}{\gamma+\frac{1}{\gamma}}\left(f\left(\frac{\varepsilon_t}{\gamma}\right)I_{[0,\infty)} + f(\gamma\varepsilon_t)I_{(-\infty,0]}\right), \tag{15}$$

where the parameter $\gamma \in (0,\infty)$ controls the skewness. Thus, if $\gamma = 1$ then $p(\varepsilon_t|\gamma)$ is symmetric, if $\gamma > 1$ then $p(\varepsilon_t|\gamma)$ is positively skewed, and if $\gamma < 1$ then $p(\varepsilon_t|\gamma)$ is negatively skewed. We considered three cases: i) the skew normal, when $f(.)$ is the standardized normal density, ii) the skew t-Student, when $f(.)$ is a t-Student with $\nu = 5$, and finally iii) the skew GED, when $f(.)$ is a GED with $\nu = 1.5$. In all cases the γ parameter was equal to 0.8 to obtain slightly asymmetric distributions for the returns. The simulated distributions are standardized as well to impose mean zero and unit variance.

In the simulation, for Gaussian disturbances, the values of parameters ϕ, σ_ω and β are chosen to be the same as those considered by Jacquier, Polson & Rossi (1994). In total, we have nine parameter combinations which represent three levels of signal-to-total (ST) ratio for the log-squared returns. From (3), the ST is defined as $ST = Var(h_t)/Var(y_t) = Var(h_t)/[Var(h_t) + Var(\eta_t)]$. Thus, in the setup, the values of ST are approximately equal to 0.32, 0.12 and 0.019. For heavy-tailed and asymmetric distributions, we choose parameters considered by Breidt & Carriquiry (1996) as those which describe the *more realistic scenarios* of Jacquier et al. (1994). Thus, the ST values are around 0.11 for heavy-tailed distributions, and around 0.134 for asymmetric distributions.

Despite the availability of R codes for estimation provided by the package `astsa`, see https://www.stat.pitt.edu/stoffer/tsa4/ and Shumway & Stoffer (2006, Chapter 6, Section 10), we developed MATLAB codes because the estimation convergence is faster compared with R. In the optimization procedure, we considered as initial parameter values: $\phi = 0.95$, $\sigma_\omega = 0.5$, and β equal to the sample mean of returns transformed by Equation (12).

The performance of the method is assessed in terms of the bias and the squared root of the mean squared error (RMSE). Each simulated time series was estimated by the method of Shumway and Stoffer using two and three terms in the mixture (4), and with and without the Fuller transformation (12), considering $c = 0.005$.

2.2 Results

In Table 1 we present the results when ε_t follows a standardized normal distribution. Cases 1-3, Cases 4-6 and Cases 7-9 correspond to ST values of 0.32, 0.12 and 0.019, respectively. The table shows that we obtained very good results in terms of bias for $\hat{\beta}$ and $\hat{\phi}$ for all cases. For instance, in terms of the estimates of ϕ, the highest bias was 9.3% of the true value (case 7, $m = 2$ with Fuller transformation), so the biases were small for all combinations of parameters. With respect to $\hat{\sigma}_\omega$, we obtained mixed results in terms of bias, with good ones in Cases 1-6 when using $m = 2$ with Fuller transformation, as well as in Cases 7-8 using $m = 3$ with Fuller transformation. In turn, the RMSE values of $\hat{\phi}$ and $\hat{\sigma}_\omega$ for Cases 7-9 were bigger than of those of Cases 1-6. In addition, for Cases 7-9, the RMSE values of $\hat{\sigma}_\omega$ were the lowest when $m = 3$ and using Fuller transformation. So, inclusion of a third term in the mixture expression and the Fuller transformation improved the results where they were difficult to estimate (cases with the lower ST values, around 0.01).

Tables 2 and 3 present the results for heavy-tailed and asymmetric distributions, respectively. For the cases considered, we also obtained very good results in terms of bias for $\hat{\beta}$ and $\hat{\phi}$, and overall good results in terms of bias for $\hat{\sigma}_\omega$ using $m = 2$ with Fuller transformation.

Finally, regarding the best combination of number of terms in the mixture ($m = 2$ or $m = 3$) and the use or not of the Fuller transformation, the results in Tables 2 and 3 indicate that using the Fuller transformation and $m = 2$ is the best choice, except in terms of the bias of $\hat{\beta}$ in four cases of Table 2. In addition, in terms of the bias of $\hat{\phi}$ it is better to use the Fuller transformation and $m = 3$ in almost all of the considered cases. For Gaussian perturbations (see Table 1), we obtained mixed results. For Cases 4, 5 and 6 (with ST values similar to those of Tables 2 and 3), and to a lesser extent for Cases 2 and 3, using the Fuller transformation and $m = 2$ was also the best option. However, for cases presenting the lowest ST values, the best combination was to use the Fuller transformation and $m = 3$ for Cases 7 and 9, and for Case 8 the best results were obtained using $m = 3$ (with or without the Fuller transformation). These results indicate that when the signal-to-total ratio is very small, more terms are needed to approximate the distribution of ε_t.

Table 1

Bias and root mean squared error (RMSE) of the estimates for SV models with perturbations ε_t following the standard Gaussian distribution. For each case, the lowest values of bias and RMSE are marked in bold. F refers to Fuller transformation and m indicates the number of terms in the mixture approximation.

						$\hat{\phi}$		$\hat{\sigma}_\omega$		$\hat{\beta}$	
case	ϕ	σ_ω	β	m		bias	RMSE	bias	RMSE	bias	RMSE
1	0.90	0.6750	0.0165	2		−0.0046	0.0226	0.0867	0.1168	**−0.0014**	**0.0021**
				3		**−0.0031**	**0.0219**	0.2169	0.2345	0.0019	0.0055
				2	F	−0.0034	0.0225	**0.0249**	**0.0759**	**0.0014**	0.0025
				3	F	0.0197	0.0318	0.1373	0.1660	0.0035	0.0046
2	0.95	0.4835	0.0165	2		−0.0043	0.0146	0.0813	0.1005	−0.0011	**0.0025**
				3		−0.0038	0.0146	0.1865	0.2002	0.0010	0.0062
				2	F	**−0.0034**	**0.0142**	**0.0419**	**0.0680**	**0.0009**	0.0027
				3	F	−0.0165	0.0234	0.1304	0.1476	0.0038	0.0055
3	0.98	0.3082	0.0165	2		−0.0035	0.0097	0.0629	0.0766	**−0.0002**	0.0074
				3		**−0.0031**	0.0098	0.1353	0.1457	0.0018	0.0098
				2	F	−0.0032	**0.0093**	**0.0415**	**0.0576**	0.0009	**0.0056**
				3	F	−0.0072	0.0149	0.0993	0.1118	0.0044	0.0125
4	0.90	0.3629	0.0252	2		−0.0097	0.0369	0.0447	0.0887	−0.0019	0.0024
				3		−0.0086	0.0356	0.0942	0.1254	0.0029	0.0075
				2	F	−0.0084	**0.0344**	**0.0273**	**0.0743**	**0.0009**	**0.0018**
				3	F	**−0.0040**	0.0351	0.0756	0.1187	0.0044	0.0063
5	0.95	0.2600	0.0252	2		−0.0068	0.0209	0.0408	0.0673	−0.0017	0.0026
				3		−0.0061	0.0205	0.0850	0.1039	0.0041	0.0074
				2	F	−0.0061	0.0201	**0.0295**	**0.0574**	**0.0008**	**0.0022**
				3	F	**−0.0037**	**0.0200**	0.0668	0.0913	0.0049	0.0067
6	0.98	0.1652	0.0252	2		−0.0054	0.0130	0.0330	0.0499	−0.0015	**0.0032**
				3		−0.0050	0.0128	0.0665	0.0791	0.0051	0.0083
				2	F	−0.0051	**0.0124**	**0.0265**	**0.0438**	**0.0008**	**0.0032**
				3	F	**−0.0040**	**0.0124**	0.0507	0.0663	0.0057	0.0082
7	0.90	0.1350	0.0293	2		−0.0816	0.1736	0.0506	0.1421	−0.0019	0.0023
				3		−0.0236	0.1422	**0.0092**	0.1350	0.0059	0.0110
				2	F	−0.0839	0.1726	0.0493	0.1329	**0.0005**	**0.0013**
				3	F	**−0.0138**	**0.1375**	**−0.0092**	**0.1243**	0.0113	0.0135
8	0.95	0.0964	0.0293	2		−0.0618	0.1381	0.0469	0.1161	−0.0019	0.0023
				3		**−0.0260**	**0.1015**	0.0241	0.1049	0.0054	0.0103
				2	F	−0.0584	0.1322	0.0410	0.1035	**0.0005**	**0.0014**
				3	F	−0.0267	0.1050	**0.0145**	**0.0968**	0.0108	0.0130
9	0.98	0.0614	0.0293	2		−0.0439	0.1112	0.0366	0.0908	−0.0018	0.0025
				3		−0.0238	0.0752	0.0301	0.0755	0.0049	0.0096
				2	F	−0.0413	0.1053	0.0326	0.0806	**0.0005**	**0.0018**
				3	F	**−0.0221**	**0.0746**	**0.0206**	**0.0670**	0.0102	0.0118

Table 2

Bias and root mean squared error (RMSE) of the estimates of SV models with heavy-tailed perturbations: (a) $\varepsilon_t \sim t_5$ and (b) $\varepsilon_t \sim$ GED ($\nu = 1.5$). In both cases, $\beta = 0.0252$. For each parameter combination, the lowest values of bias and RMSE are marked in bold. F refers to Fuller transformation and m indicates the number of terms in the mixture approximation.

				$\hat{\phi}$		$\hat{\sigma}_\omega$		$\hat{\beta}$	
ϕ	σ_ω	m		bias	RMSE	bias	RMSE	bias	RMSE
(a) t-Student									
0.9000	0.3629	2		−0.0121	0.0434	0.0404	0.0989	−0.0058	0.0059
		3		−0.0105	**0.0418**	0.0897	0.1362	−0.0021	0.0069
		2	F	−0.0111	**0.0418**	**0.0165**	**0.0852**	**−0.0018**	**0.0024**
		3	F	**−0.0085**	0.0465	0.0579	0.1269	−0.0020	0.0060
0.9500	0.2600	2		−0.0090	0.0254	0.0384	0.0720	−0.0057	0.0059
		3		−0.0085	0.0249	0.0833	0.1086	**−0.0019**	0.0067
		2	F	−0.0086	**0.0242**	**0.0228**	**0.0603**	−0.0020	**0.0028**
		3	F	**−0.0078**	0.0255	0.0581	0.0934	−0.0017	0.0060
0.9800	0.1652	2		−0.0067	0.0151	0.0317	0.0513	−0.0055	0.0061
		3		−0.0065	0.0152	0.0657	0.0810	**−0.0017**	0.0071
		2	F	−0.0066	**0.0148**	**0.0225**	**0.0438**	−0.0022	**0.0036**
		3	F	**−0.0064**	0.0158	0.0486	0.0698	−0.0018	0.0068
(b) GED									
0.9000	0.3629	2		−0.0117	0.0419	0.0435	0.0946	−0.0035	0.0038
		3		−0.0096	0.0397	0.0927	0.1323	0.0023	0.0068
		2	F	−0.0099	**0.0383**	**0.0210**	**0.0783**	**0.0002**	**0.0016**
		3	F	**−0.0060**	0.0403	0.0511	0.1086	0.0040	0.0067
0.9500	0.2600	2		−0.0080	0.0229	0.0400	0.0703	−0.0034	0.0039
		3		−0.0072	0.0234	0.0855	0.1084	0.0024	0.0069
		2	F	−0.0074	**0.0226**	**0.0264**	**0.0595**	**−0.0001**	**0.0021**
		3	F	**−0.0065**	0.0246	0.0562	0.0872	0.0042	0.0067
0.9800	0.1652	2		−0.0062	0.0133	0.0331	0.0515	−0.0033	0.0042
		3		−0.0057	0.0134	0.0687	0.0829	0.0034	0.0069
		2	F	−0.0059	**0.0129**	**0.0255**	**0.0442**	**−0.0002**	**0.0031**
		3	F	**−0.0052**	0.0136	0.0467	0.0650	0.0045	0.0076

Table 3

Bias and root mean squared error (RMSE) of the estimates of SV models with skew perturbations: (a) $\varepsilon_t \sim$ Skew normal ($\gamma = 0.8$), (b) $\varepsilon_t \sim$ Skew t-Student ($\gamma = 0.8$, $\nu = 5$) and (c) $\varepsilon_t \sim$ Skew GED ($\gamma = 0.8$, $\nu = 1.5$). In all cases, $\beta = 0.0252$. For each parameter combination, the lowest values of bias and RMSE are marked in bold. F refers to Fuller transformation and m indicates the number of terms in the mixture approximation.

				$\hat{\phi}$		$\hat{\sigma}_\omega$		$\hat{\beta}$	
ϕ	σ_ω	m		bias	RMSE	bias	RMSE	bias	RMSE
(a) skew normal									
0.9000	0.3629	2		−0.0084	0.0357	0.0420	0.0878	−0.0020	0.0025
		3		−0.0072	0.0347	0.0907	0.1236	0.0018	0.0076
		2	F	−0.0069	**0.0331**	**0.0244**	**0.0736**	**0.0006**	**0.0017**
		3	F	**−0.0040**	0.0351	0.0756	0.1187	0.0044	0.0063
0.9500	0.2600	2		−0.0064	0.0206	0.0405	0.0671	−0.0018	0.0027
		3		−0.0055	0.0199	0.0829	0.1019	0.0027	0.0074
		2	F	−0.0056	**0.0195**	**0.0286**	**0.0569**	**0.0005**	**0.0021**
		3	F	**−0.0037**	0.0200	0.0668	0.0913	0.0049	0.0067
0.9800	0.1652	2		−0.0053	0.0124	0.0333	0.0494	−0.0016	0.0032
		3		−0.0047	0.0121	0.0651	0.0770	0.0039	0.0086
		2	F	−0.0050	**0.0119**	**0.0264**	**0.0430**	**0.0005**	**0.0031**
		3	F	**−0.0040**	0.0124	0.0507	0.0663	0.0057	0.0082
(b) skew t-Student									
0.9000	0.3629	2		−0.0104	0.0403	0.0386	0.0956	−0.0060	0.0062
		3		−0.0103	0.0423	0.0889	0.1332	−0.0033	0.0072
		2	F	−0.0093	**0.0389**	**0.0153**	**0.0828**	**−0.0025**	**0.0029**
		3	F	**−0.0086**	0.0465	0.0583	0.1267	−0.0020	0.0060
0.9500	0.2600	2		−0.0074	0.0228	0.0368	0.0696	−0.0060	0.0062
		3		−0.0071	0.0229	0.0790	0.1040	−0.0034	0.0073
		2	F	**−0.0070**	**0.0225**	**0.0207**	**0.0594**	−0.0028	**0.0034**
		3	F	−0.0078	0.0255	0.0582	0.0934	**−0.0017**	0.0060
0.9800	0.1652	2		−0.0058	0.0136	0.0305	0.0504	−0.0059	0.0063
		3		**−0.0054**	**0.0134**	0.0611	0.0768	−0.0033	0.0075
		2	F	−0.0057	0.0135	**0.0206**	**0.0430**	−0.0030	**0.0040**
		3	F	−0.0064	0.0158	0.0486	0.0698	**−0.0017**	0.0068
(c) skew GED									
0.9000	0.3629	2		−0.0100	0.0394	0.0396	0.0932	−0.0039	0.0041
		3		−0.0090	0.0395	0.0938	0.1335	−0.0006	0.0078
		2	F	−0.0083	**0.0366**	**0.0192**	**0.0798**	**−0.0005**	**0.0017**
		3	F	**−0.0060**	0.0403	0.0511	0.1086	0.0040	0.0067
0.9500	0.2600	2		−0.0071	0.0223	0.0375	0.0690	−0.0038	0.0042
		3		**−0.0064**	0.0217	0.0855	0.1078	**−0.0004**	0.0079
		2	F	**−0.0064**	**0.0214**	**0.0244**	**0.0591**	−0.0007	**0.0021**
		3	F	**−0.0064**	0.0246	0.0562	0.0872	0.0042	0.0067
0.9800	0.1652	2		−0.0057	0.0133	0.0312	0.0502	−0.0036	0.0045
		3		**−0.0051**	**0.0129**	0.0676	0.0816	0.0010	0.0081
		2	F	−0.0054	**0.0129**	**0.0237**	**0.0437**	**−0.0008**	**0.0030**
		3	F	−0.0052	0.0136	0.0467	0.0650	0.0045	0.0076

3. VaR forecasting

In this section, we assess the method of Shumway and Stoffer empirically, through the estimation of one-step-ahead VaR forecasts of the S&P 500 returns. Our sample spans from January 6, 2003 to July 8, 2019, a total of 4,152 days. The data were obtained from Bloomberg, and all holidays were excluded to avoid zero returns.

To obtain the one-step-ahead VaR predictions we proceeded as follows. Once the parameters are estimated, based on the Kalman Filter presented by Shumway & Stoffer (2006, Chapter 6, Section 10) and reproduced in the Introduction, it is straightforward to obtain the one-step-ahead forecast of h_t, denoted by $\hat{h}_{T+1|T}$, where T is the end of the sample. Then, the one-step-ahead volatility forecast is given by $\hat{\sigma}_{T+1|T} = \hat{\beta}\exp(0.5\hat{h}_{T+1|T})$. Therefore, from Equation (1) an estimator of the $100(1-\alpha)\%$ VaR is given by $\hat{q}_\alpha\hat{\sigma}_{T+1|T}$, where $\hat{q}_\alpha$ is an estimator of the α-quantile of ε_t.

Finding $\hat{q}_\alpha$ from the data poses a challenging problem because the method applied does not specify a parametric model for $\eta_t = \log(\varepsilon_t^2)$; instead a mixture model is proposed for η_t. To overcome this difficulty, we proceed as follows. First, for $t = 1,\ldots,T$ we calculated the smoothed volatilities $\hat{\sigma}_{t|T} = \hat{\beta}\exp(0.5\hat{h}_{t|T})$ where $\hat{h}_{t|T}$ are obtained via a Kalman smoother. Second, we estimated ε_t by $e_t = r_t/\hat{\sigma}_{t|T}$. Third, $\hat{q}_\alpha$ was calculated as the α-quantile of the standardized empirical distribution of $e_1,\ldots,e_T$.

The returns of the period from February 23, 2012 to July 8, 2019 were used in a backtesting exercise to evaluate the $100(1-\alpha)\%$-VaR forecasts using a rolling-window sample of 2,500 returns. Thus, for each of these dates, a VaR forecast was obtained using the previous 2,500 returns. For each window, both parameters and volatilities were estimated again. Like the Monte Carlo experiment, our objective in this section is to compare the VaR forecast using $m = 2$ and $m = 3$, and evaluate if the Fuller transformation is convenient. However, we stress that the Fuller transformation was applied only to estimate the parameters of the SV model and the forecasts were obtained using the original data.

As an illustration, in Table 4 we present the parameter estimates using $m = 2$ and $m = 3$, as well as compare them with and without the use of the Fuller transformation, over the period from January 6, 2003 to February 22, 2012 (2,500 observations). This is an important period because it includes the financial crisis of October 2008 and some periods of high volatility, such as the beginning of August 2011 (the period of the downgrade of US bond rating). Table 4 shows that the estimates of ϕ are around 0.98 (strong persistence), and the values of $\hat{\sigma}_\omega$ are around 0.2. Therefore, the estimates of the S&P 500 are close of the parameters of Case 6 in Table 1.

Table 4

Results of SV estimation of S&P 500 returns of the period from 06-Jan-2003 to 22-Feb-2012 (sample size of 2,500 observations). The errors, obtained from the Hessian matrix, are in parentheses.

	without Fuller transformation		with Fuller transformation	
	$m = 2$	$m = 3$	$m = 2$	$m = 3$
ϕ	0.988 (0.005)	0.983 (0.005)	0.987 (0.005)	0.986 (0.005)
σ_ω	0.187 (0.028)	0.251 (0.034)	0.181 (0.025)	0.223 (0.029)
α	−9.127 (0.245)	−9.767 (0.234)	−9.456 (0.284)	−8.956 (0.230)
σ_1	1.128 (0.038)	0.727 (0.047)	0.997 (0.035)	0.714 (0.038)
μ_2	−2.575 (0.090)	−2.059 (0.082)	−2.851 (0.080)	−4.140 (0.092)
σ_2	2.636 (0.074)	1.066 (0.064)	1.644 (0.050)	1.347 (0.058)
μ_3		−3.939 (0.177)		−1.758 (0.070)
σ_3		2.809 (0.111)		0.849 (0.062)

The quality of VaR forecasts was assessed by analyzing the proportion of *violations*, where a violation occurs when the loss is bigger than the forecast VaR. We applied the Kupiec test (Kupiec, 1995) to check if the proportion of violations was close to the theoretical value, the Christoffersen (1998) test to check conditional coverage, and the test proposed by Christoffersen & Pelletier (2004) to check if the violations were independent. The results are presented in Table 5.

Table 5

Results of backtesting analysis, presenting the proportion of violations of 100(1-α)%-VaR, and the p-values of the Kupiec, Christoffersen, and Christoffersen and Pelletier tests. The Fuller transformation was only applied to estimate the parameters of the SV model.

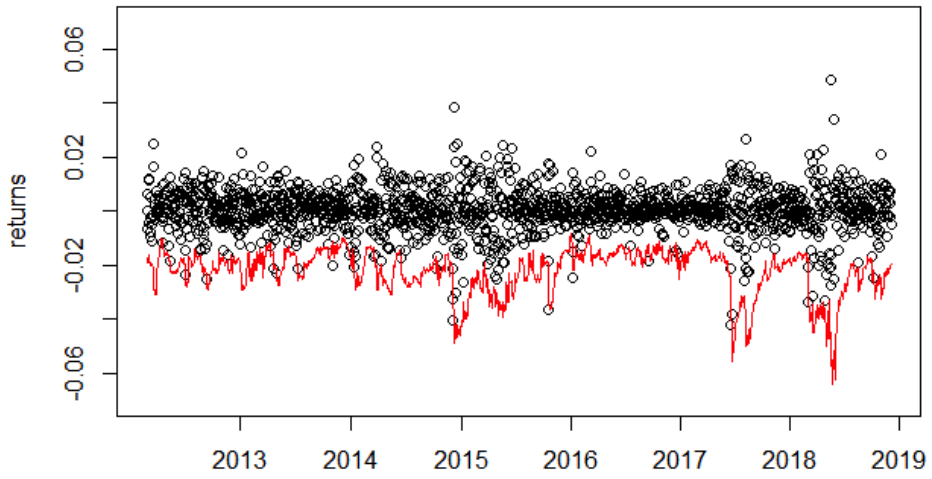
	without Fuller transformation		with Fuller transformation	
	$m = 2$	$m = 3$	$m = 2$	$m = 3$
		α = 0.05		
proportion	0.065	0.048	0.049	0.033
Kupiec	0.006	0.768	0.856	0.001
Christoffersen	0.012	0.142	0.002	0.002
Christoffersen and Pelletier	0.290	0.075	0.001	0.303
		α = 0.025		
proportion	0.042	0.034	0.041	0.024
Kupiec	0.000	0.028	0.000	0.837
Christoffersen	0.000	0.034	0.000	0.226
Christoffersen and Pelletier	0.641	0.393	0.005	0.151
		α = 0.01		
proportion	0.028	0.026	0.030	0.012
Kupiec	0.000	0.000	0.000	0.549
Christoffersen	0.000	0.000	0.000	0.386
Christoffersen and Pelletier	0.063	0.026	0.000	0.427

Taking into account both the percentage of violations and the p-values of the tests, the following conclusions emerge from Table 5. For 95% VaRs, the best results are obtained using $m = 3$ without the Fuller transformation (good results) and the second-best results, in terms of nominal value, are obtained using $m = 2$ and the Fuller transformation. For 97.5% and 99% VaRs, the best results are obtained using $m = 3$ and Fuller transformation. These results are very good in terms of both percentage of violations and performance of the tests. For 97.5% VaRs the second-best results are obtained using $m = 3$ without Fuller transformation, considered good results even though there is overestimation. For 99% VaRs, except when using $m = 3$ and Fuller transformation, the results are not good, indicating the importance of both the transformation and the number of terms in the mixture.

In Figure 1 we present the graph of the returns in the backtesting period and the 99%-VaR forecasts calculated using $m = 3$ and the Fuller transformation. These forecasts mimic very well the movements of the volatility and they are not far below the losses. Therefore, the behavior of the VaR forecasts is very good.

Figure 1

Backtesting results for S&P returns in black dots and predicted 99%-VaR, with Fuller transformation and $m = 3$ in red line.



4. Conclusions

Monte Carlo experiments revealed that the method for estimating SV models proposed by Shumway & Stoffer (2006, Chapter 6, Section 10) presents good performance overall in terms of point estimates, considering several distributions for the perturbations (ε_t) in the observation equation. However, the estimation of $\hat{\sigma}_\omega$ in terms of bias is not good in all cases, because of the existence of high bias values. Overall, the best results were obtained using $m = 2$ terms in the mixture and using the Fuller transformation, except for the cases of very low signal-to-total ratio, where $m = 3$ terms are needed. On the other hand, the backtesting analysis of the S&P 500 returns revealed that in terms of one-step-ahead 97.5% and 99% VaR forecasts, using $m = 3$ and Fuller transformation provides very good results. Therefore, prudent advice to practitioners who want to use the method discussed in this paper is to use Fuller transformation with both $m = 2$ and $m = 3$ terms in the mixture, and decide between the two values of m based on backtesting considerations.

5. Acknowledgments

The first author acknowledges financial support from CAPES Ph.D. scholarship. The second author acknowledges financial support from the São Paulo Research Foundation (FAPESP) grant 2018/04654-9 and the support of the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS). The authors thank an anonymous referee for comments and valuable suggestions that led to an improvement in the paper.

References

Bai, X., Russell, J. R. & Tiao, G. C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114**(2): 349 - 360.

Breidt, F.J. & Carriquiry, A.L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In *Modelling and Prediction: Honoring Seymour Geisser*, 228–247. Lee, J.C, Johnson, W.O. and Zellner, A. (eds). Springer-Verlag, New York.

Broto, C. & Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys* **18**: 613-649.

Christoffersen, P. (1998). Evaluating Interval Forecasts. *Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance* **39**(4): 841-862.

Christoffersen, P. & Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Empirical Finance* **2**: 84–108.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987-1007.

Fernandez, C. & Steel, M.F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**: 359-371.

Fuller, W. (1996). *Introduction to Statistical Time Series*, Second Edition. Wiley, New York.

Harvey, A.C., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies* **61**: 247-264.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* **12**: 371-389.

Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies* **65**: 361-393.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3**: 73-84.

Liesenfeld, R. & Jung, R.C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* **15**: 137-160.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**(2): 347-370.

Shephard, N. & Andersen, T.G. (2009). Stochastic volatility: origins and overview. In *Handbook of Financial Time Series*, 233–254. Andersen, T.G, Davis, R.A., Kreβ, J.-P., Mikosch, T. (eds), Springer Verlag, Berlin-Heidelberg.

Shumway, R. H. & Stoffer, D. S. (2006). *Time Series Analysis and Its Applications*. Springer Verlag, New York.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes – A study of daily sugar prices, 1961-79. In Anderson, O. D., editor, *Time Series Analysis: Theory and Practice*, Volume 1, pages 203-226. Elsevier/North-Holland, New York.

Taylor, S.J. (1986). *Modelling Financial Time Series*. Wiley, New York.

Zevallos, M., Gasco, L. & Ehlers, R. (2017). Riemann manifold Langevin methods on stochastic volatility estimation, *Communications in Statistics – Simulation and Computation* 46(10): 7942-7956.